

Visual Analytics: Scope and Challenges

Daniel A. Keim, Florian Mansmann, Jörn Schneidewind, Jim Thomas, and
Hartmut Ziegler

University of Konstanz,
{keim, mansmann, schneide, ziegler}@informatik.uni-konstanz.de,
Website: <http://infovis.uni-konstanz.de>

Pacific Northwest National Laboratory,
National Visualization and Analytics Center (NVAC),
nvac@pnl.gov,
Website: <http://nvac.pnl.gov>

Abstract. In today's applications data is produced at unprecedented rates. While the capacity to collect and store new data rapidly grows, the ability to analyze these data volumes increases at much lower rates. This gap leads to new challenges in the analysis process, since analysts, decision makers, engineers, or emergency response teams depend on information hidden in the data. The emerging field of visual analytics focuses on handling these massive, heterogeneous, and dynamic volumes of information by integrating human judgement by means of visual representations and interaction techniques in the analysis process. Furthermore, it is the combination of related research areas including visualization, data mining, and statistics that turns visual analytics into a promising field of research. This paper aims at providing an overview of visual analytics, its scope and concepts, addresses the most important research challenges and presents use cases from a wide variety of application scenarios.

1 Introduction

The information overload is a well-known phenomenon of the information age, since due to the progress in computer power and storage capacity over the last decades, data is produced at an incredible rate, and our ability to collect and store these data is increasing at a faster rate than our ability to analyze it. But, the analysis of these massive, typically messy and inconsistent, volumes of data is crucial in many application domains. For decision makers, analysts or emergency response teams it is an essential task to rapidly extract relevant information from the flood of data. Today, a selected number of software tools is employed to help analysts to organize their information, generate overviews and explore the information space in order to extract potentially useful information. Most of these data analysis systems still rely on interaction metaphors developed more than a decade ago and it is questionable whether they are able to meet the demands of the ever-increasing mass of information. In fact, huge investments in time and money are often lost, because we still lack the possibilities to

properly interact with the databases. Visual analytics aims at bridging this gap by employing more intelligent means in the analysis process. The basic idea of visual analytics is to visually represent the information, allowing the human to directly interact with the information, to gain insight, to draw conclusions, and to ultimately make better decisions. The visual representation of the information reduces complex cognitive work needed to perform certain tasks. People may use visual analytics tools and techniques to synthesize information and derive insight from massive, dynamic, and often conflicting data by providing timely, defensible, and understandable assessments.

The goal of visual analytics research is to turn the information overload into an opportunity. Decision-makers should be enabled to examine this massive, multi-dimensional, multi-source, time-varying information stream to make effective decisions in time-critical situations. For informed decisions, it is indispensable to include humans in the data analysis process to combine flexibility, creativity, and background knowledge with the enormous storage capacity and the computational power of today's computers. The specific advantage of visual analytics is that decision makers may focus their full cognitive and perceptual capabilities on the analytical process, while allowing them to apply advanced computational capabilities to augment the discovery process. This paper gives an overview on visual analytics, and discusses the most important research challenges in this field. Real world application examples are presented that show how visual analytics can help to turn information overload as generated by today's applications into useful information.

The rest of the paper is organized as follows: section 2 defines visual analytics and discusses its scope. The visual analytics process is formalized in section 3. Section 4 covers the 10 most important application challenges in the field and presents some approaches addressing these problems. It is followed by the 10 most important technical challenges in section 5. Finally, section 6 concludes our work and gives a short outlook of the future of visual analytics.

2 Scope of Visual Analytics

In general, *visual analytics* can be described as “the science of analytical reasoning facilitated by interactive visual interfaces” [1]. To be more precise, visual analytics is an iterative process that involves information gathering, data preprocessing, knowledge representation, interaction and decision making. The ultimate goal is to gain insight in the problem at hand which is described by vast amounts of scientific, forensic or business data from heterogeneous sources. To reach this goal, visual analytics combines the strengths of machines with those of humans. On the one hand, methods from knowledge discovery in databases (KDD), statistics and mathematics are the driving force on the automatic analysis side, while on the other hand human capabilities to perceive, relate and conclude turn visual analytics into a very promising field of research.

Historically, visual analytics has evolved out of the fields of information and scientific visualization. According to Colin Ware, the term visualization is mean-

while understood as “a graphical representation of data or concepts” [2], while the term was formerly applied to form a mental image. Nowadays fast computers and sophisticated output devices create meaningful visualizations and allow us not only to mentally visualize data and concepts, but also to see and explore an exact representation of the data under consideration on a computer screen. However, the transformation of data into meaningful visualizations is not a trivial task that will automatically improve through steadily growing computational resources. Very often, there are many different ways to represent the data under consideration and it is unclear which representation is the best one. State-of-the-art concepts of representation, perception, interaction and decision-making need to be applied and extended to be suitable for visual data analysis.

The fields of information and scientific visualization deal with visual representations of data. The main difference among the two is that scientific visualization examines potentially huge amounts of scientific data obtained from sensors, simulations or laboratory tests. Typical scientific visualization applications are flow visualization, volume rendering, and slicing techniques for medical illustrations. In most cases, some aspects of the data can be directly mapped onto geographic coordinates or into virtual 3D environments. We define Information visualization more generally as the communication of abstract data relevant in terms of action through the use of interactive interfaces. There are three major goals of visualization, namely a) presentation, b) confirmatory analysis, and c) exploratory analysis. For presentation purposes, the facts to be presented are fixed a priori, and the choice of the appropriate presentation technique depends largely on the user. The aim is to efficiently and effectively communicate the results of an analysis. For confirmatory analysis, one or more hypotheses about the data serve as a starting point. The process can be described as a goal-oriented examination of these hypotheses. As a result, visualization either confirms these hypotheses or rejects them. *Exploratory data analysis* as the process of searching and analyzing databases to find implicit but potentially useful information, is a difficult task. At the beginning, the analyst has no hypothesis about the data. According to John Tuckey, tools as well as understanding are needed [3] for the interactive and usually undirected search for structures and trends.

Visual analytics is more than only visualization. It can rather be seen as an integral approach combining visualization, human factors and data analysis. Figure 1 illustrates the detailed scope of visual analytics. Concerning the field of visualization, visual analytics integrates methodology from information analytics, geospatial analytics, and scientific analytics. Especially human factors (e.g., interaction, cognition, perception, collaboration, presentation, and dissemination) play a key role in the communication between human and computer, as well as in the decision-making process. In this context, *production* is defined as the creation of materials that summarize the results of an analytical effort, *presentation* as the packaging of those materials in a way that helps the audience understand the analytical results in context using terms that are meaningful to them, and *dissemination* as the process of sharing that information with the intended audience [4]. In matters of data analysis, visual analytics further-

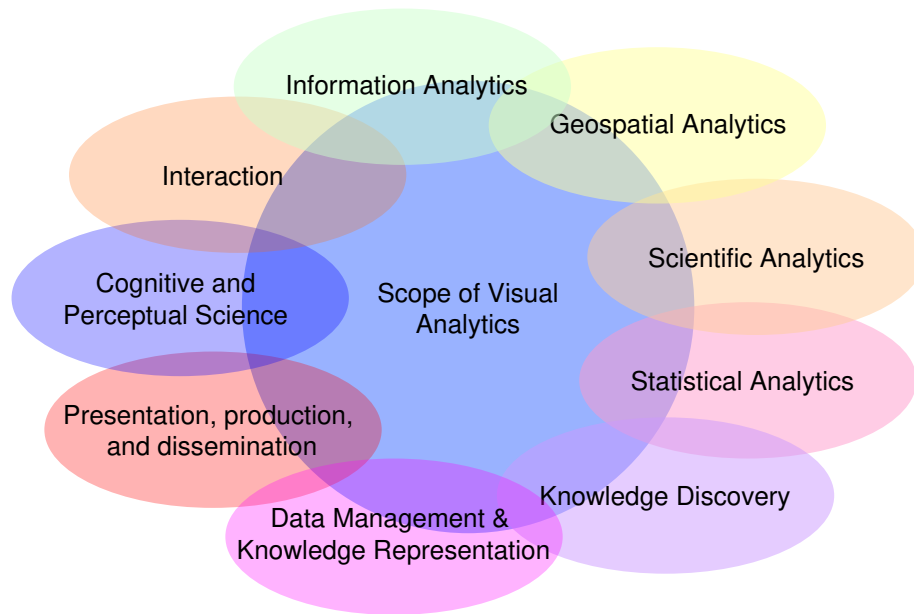


Fig. 1. The Scope of Visual Analytics

more profits from methodologies developed in the fields of data management & knowledge representation, knowledge discovery and statistical analytics. Note that visual analytics, is not likely to become a separate field of study [5], but its influence will spread over the research areas it comprises.

According to Jarke J. van Wijk, “visualization is not ‘good’ by definition, developers of new methods have to make clear why the information sought cannot be extracted automatically” [6]. From this statement, we immediately see the need for the visual analytics approach using automatic methods from statistics, mathematics and knowledge discovery in databases (KDD) wherever they are applicable. Visualization is used as a means to efficiently communicate and explore the information space when automatic methods fail. In this context, human background knowledge, intuition and decision-making either cannot be automated or serve as input for the future development of automated processes.

Overlooking a large information space is a typical visual analytics problem. In many cases, the information at hand is conflicting and needs to be integrated from heterogeneous data sources. Moreover, the system lacks knowledge that is still hidden in the expert’s mind. By applying analytical reasoning, hypotheses about the data can be either affirmed or discarded and eventually lead to a better understanding of the data, thus supporting the analyst in his task to gain insight. Contrary to this, a well-defined problem where the optimum or a good estimation can be calculated by non-interactive analytical means would rather not be described as a visual analytics problem. In such a scenario, the non-interactive

analysis should be clearly preferred due to efficiency reasons. Likewise, visualization problems not involving methods for automatic data analysis do not fall into the field of visual analytics.

The fields of visualization and visual analytics both build upon methods from scientific analytics, geospatial analytics and information analytics. They both profit from knowledge out of the field of interaction as well as cognitive and perceptual science. They do differentiate in so far as visual analytics furthermore integrates methodology from the fields of statistical analytics, knowledge discovery, data management & knowledge representation and presentation, production & dissemination.

3 Visual Analytics Process

In this section we provide a formal description of the visual analytics process. As described in the last section the input for the data sets used in the visual analytics process are heterogeneous data sources (i.e., the internet, newspapers, books, scientific experiments, expert systems). From these rich sources, the data sets $S = S_1, \dots, S_m$ are chosen, whereas each $S_i, i \in (1, \dots, m)$ consists of attributes A_{i1}, \dots, A_{ik} . The goal or output of the process is insight I . Insight is either directly obtained from the set of created visualizations V or through confirmation of hypotheses H as the results of automated analysis methods. We illustrated this formalization of the visual analytics process in Figure 2. Arrows represent the transitions from one set to another one.

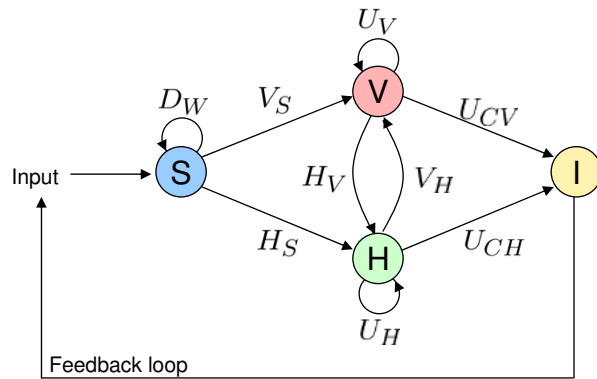


Fig. 2. Visual Analytics Process

More formal the visual analytics process is a transformation $F : S \rightarrow I$, whereas F is a concatenation of functions $f \in \{D_W, V_X, H_Y, U_Z\}$ defined as follows:

D_W describes the basic data pre-processing functionality with $D_W : S \rightarrow S$ and $W \in \{T, C, SL, I\}$ including data transformation functions D_T , data cleaning functions D_C , data selection functions D_{SL} and data integration functions D_I that are needed to make analysis functions applicable to the data set.

$V_W, W \in \{S, H\}$ symbolizes the visualization functions, which are either functions visualizing data $V_S : S \rightarrow V$ or functions visualizing hypotheses $V_H : H \rightarrow V$.

$H_Y, Y \in \{S, V\}$ represents the hypothesis generation process. We distinguish between functions that generate hypotheses from data $H_S : S \rightarrow H$ and functions that generate hypotheses from visualizations $H_V : V \rightarrow H$.

Moreover, user interactions $U_Z, Z \in \{V, H, CV, CH\}$ are an integral part of the visual analytics process. User interactions can either effect only visualizations $U_V : V \rightarrow V$ (i.e., selecting or zooming), or can effect only hypotheses $U_H : H \rightarrow H$ by generating a new hypotheses from given ones. Furthermore, insight can be concluded from visualizations $U_{CV} : V \rightarrow I$ or from hypothesis $U_{CH} : H \rightarrow I$

The typical data pre-processing applying data cleaning, data integration and data transformation functions is defined as $D_P = D_T(D_I(D_C(S_1, \dots, S_n)))$. After the pre-processing step either automated analysis methods $H_S = \{f_{s1}, \dots, f_{sq}\}$ (i.e., statistics, data mining, etc.) or visualization methods $V_S : S \rightarrow V, V_S = \{f_{v1}, \dots, f_{vs}\}$ are applied to the data, in order to reveal patterns as shown in Figure 2.

The application of visualization methods can hereby directly provide insight to the user, described by U_{CV} ; the same applies to automatic analysis methods U_{CH} . However, most application scenarios may require user interaction to refine parameters in the analysis process and to steer the visualization process. This means that after having obtained initial results from either the automatic analysis step or the visualization step, the user may refine the achieved results by applying another data analysis step, expressed by U_V and U_H . Furthermore visualization methods can be applied to the results of the automated analysis step to transform a hypotheses into a visual representation V_H or the findings extracted from visualizations may be validated through an data analysis step to generated a hypotheses H_V . F(S) is rather an iterative process than a single application of each provided function, as indicated by the feedback loop in Figure 2. The user may refine input parameters or focus on different parts of the data in order to validate generated hypotheses or extracted insight.

We take a visual analytics application for monitoring network security as an example. Within the network system, four sensors measure the network traffic resulting in four data sets S_1, \dots, S_4 . While preprocessing, the data is cleaned from missing values and unnecessary data using the data cleaning function d_c , integrated using d_i (each measurement system stores data slightly different), and transformed in a format suitable for our analysis using d_t . We now select UDP and TCP traffic for our analysis with the function d_s , resulting in $S' = d_s(d_t(d_i(d_c(S_1, \dots, S_4))))$. For further analysis, we apply a data mining algorithm h_s to search for security incidents within the traffic generating a hypothesis $h' = h_s(S')$. To better understand this hypothesis, we visualize it using

the function v_h : $v' = v_h(h')$. Interactive adjustment of the parameters results in $v'' = u_v(v')$, revealing a correlation of the incidents from two specific source networks. By applying the function h_v , we obtain a distribution of networks where similar incidents took place $h'' = h_v(v'')$. This leads to the insight that a specific network worm tries to communicate with our network from 25 source networks $i' = u_{ch}(h'')$. Repeating the same process at a later date by using the feedback loop reveals a much higher spread of the virus, emphasizing the need to take countermeasures.

Unlike described in the information seeking mantra (“overview first, zoom/filter, details on demand”) [7], the visual analytics process comprises the application of automatic analysis methods before and after the interactive visual representation is used like demonstrated in the example. This is primarily due to the fact that current and especially future data sets are complex on the one hand and too large to be visualized straightforward on the other hand. Therefore, we present the visual analytics mantra:

*“Analyse First -
Show the Important -
Zoom, Filter and Analyse Further -
Details on Demand”*

4 Application Challenges

For the advancement of the research field of visual analytics several application and technical challenges need to be mastered. In this section, we present the ten most significant application challenges and discuss them in the context of research projects trying to solve the challenges. Both the application (this section), as well as the technical challenges (next section) were identified by the panel discussion on the Workshop on Visual Analytics in 2005 [8].

4.1 Physics and Astronomy

One major field in the area of visual analytics covers physics and astronomy, including applications like flow visualization, fluid dynamics, molecular dynamics, nuclear science and astrophysics, to name just a few of them.

Especially the research field of astrophysics offers a wide variety of usage scenarios for visual analytics. Never before in history scientists had the ability to capture so much information about the universe. Massive volumes of unstructured data, originating from different directions of the orbit and covering the whole frequency spectrum, form continuous streams of terabytes of data that can be recorded and analysed. The amount of data is so high that it far exceeds the ability of humans to consider it all. By common data analysis techniques like knowledge discovery, astronomers can find new phenomena, relationships and useful knowledge about the universe, but although a lot of the data only consists of noise, a visual analytics approach can help separating relevant data

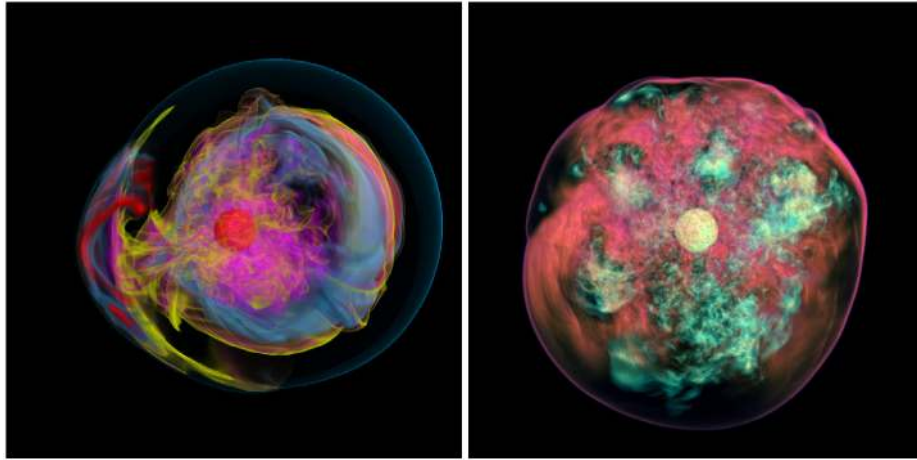


Fig. 3. A visual approach to illustrate the complex relationships within a Supernova (© 2005 IEEE) [9]. The 3D simulation processes tens of terabytes of data (turbulence, rotation, radiation, magnetic fields, gravitational forces) to generate a visual output that can then be analysed to discover further insights.

from noise and help identifying unexpected phenomena inside the massive and dynamic data streams. One celebrated example is the Sloan Digital Sky Survey [10] and the COMPLETE project [11], generating terabytes of astrophysics data each day, or the Large Hadron Collider (LHC) at CERN which generates a volume of 1 petabyte of data per year.

One example for a visual analytics application is the simulation of a Supernova. The SciDAC program has brought together tremendous scientific expertise and computing resources within the Terascale Supernova Initiative (TSI) project to realize the promise of terascale computing for attempting to answer some of the involved questions [9]. A complete understanding of core collapse supernovae requires 3D simulations of the turbulence, rotation, radiation, magnetic fields and gravitational forces, producing tens of terabytes of data per simulation. As an examination of this massive amount of data in a numeric format would simply exceed human capabilities and would therefore not give an insight into the processes involved, a visual approach (see Fig. 3) can help analyzing these processes on a higher aggregated level in order to draw conclusions and extract knowledge from it.

4.2 Business

Another major field in the area of visual analytics covers business applications. The financial market with its thousands of different stocks, bonds, futures, commodities, market indices and currencies generates a lot of data every second, which accumulates to high data volumes throughout the years. The main chal-

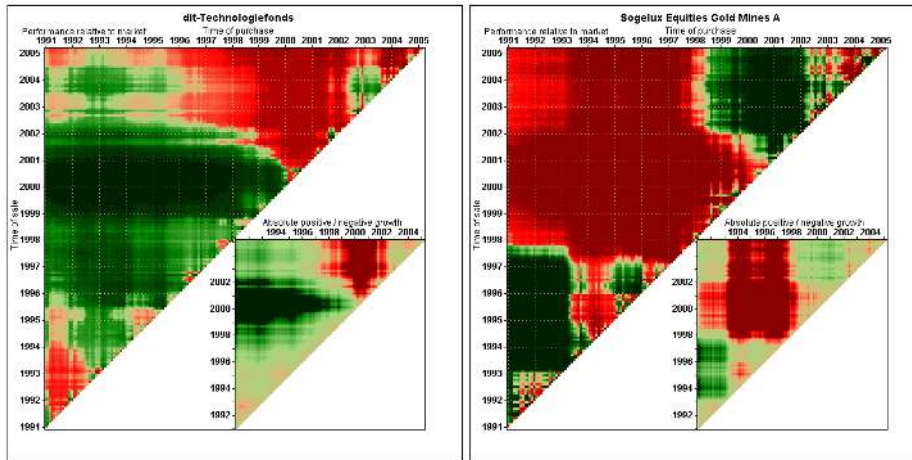


Fig. 4. Visual analysis of financial data with the FinDEx system [12]. The growth rates for time intervals are triangulated in order to visualize all possible time frames. The small triangle represents the absolute performance of one stock, the big triangle represents the performance of one stock compared to the whole market.

lenge in this area lies in analyzing the data under multiple perspectives and assumptions to understand historical and current situations, and then monitoring the market to forecast trends and to identify recurring situations. Visual analytics applications can help analysts obtaining insights and understanding into previous stock market development, as well as supporting the decision making progress by monitoring the stock market in real-time in order to take necessary actions for a competitive advantage, with powerful means that reach far beyond the numeric technical chart analysis indicators or traditional line charts. One popular application in this field is the well-known Smartmoney [13], which gives an instant visual overview of the development of the stock market in particular sectors for a user-definable time frame. A new application in this field is the FinDEx system [12] (see Fig. 4), which allows a visual comparison of a fund's performance to the whole market for all possible time intervals at one glance.

4.3 Environmental Monitoring

Monitoring climate and weather is also a domain which involves huge amounts of data collected throughout the world or from satellites in short time intervals, easily accumulating to terabytes per day. Applications in this domain most often do not only visualize snapshots of a current situation, but also have to generate sequences of previous developments and forecasts for the future in order to analyse certain phenomena and to identify the factors responsible for a development, thus enabling the decision maker to take necessary countermeasures (like the global reduction of carbon dioxide emissions in order to reduce global

warming). The applications for climate modeling and climate visualization can cover all possible time intervals, from daily weather forecasts which operate in rather short time frames of several days, to more complex visualizations of climate changes that can expand to thousands of years. A visual approach can easily help to interpret these massive amounts of data and to gain insight into the dependencies of climate factors and climate change scenarios that would otherwise not be easily identified. Besides weather forecasts, existing applications for instance visualize the global warming, melting of the poles, the stratospheric ozone depletion, hurricane warnings or oceanography, to name just a few.

4.4 Disaster and Emergency Management

Despite the slowly arising environmental changes like global warming that have been mentioned above, environmental or other disasters can face us as sudden major catastrophes. In the domain of emergency management, visual analytics can help determining the on-going progress of an emergency and can help identifying the next countermeasures (construction of physical countermeasures or evacuation of the population) that must be taken to limit the damage. Such scenarios can include natural or meteorological catastrophes like flood or waves, volcanos, storm, fire or epidemic growth of diseases (bird flu), but also human-made technological catastrophes like industrial accidents, transport accidents or pollution. Depending on the particular case, visual analytics can help to determine the amount of damage, to identify objectives, to assign priorities, and to provide effective coordination for various organizations for more efficient help in the disaster zone.

4.5 Security

Visual analytics for security is an important research topic and is strongly supported by the U.S. government. The application field in this sector is wide, ranging from terrorism informatics over border protection to network security. In these fields, the challenges lie in getting all the information together and linking numerous incidents to find correlations.

A demonstrative example of work in the field is the situational awareness display *VisAware* [14] which is built upon the w^3 premise, assuming that every incident has at least the three attributes what, when, and where (see Fig. 5). In this display, the location attribute is placed on a map, the time attribute indicated on concentric circles around this map, and the classification of the incident is mapped to the angle around the circle. For each incident, the attributes are linked through lines. Other examples in the field are [15] and [16].

4.6 Software Analytics

Visual software analytics has become a popular research area, and as modern software packages often consist of millions of code lines it can support a faster understanding of the structure of a software package with its dependencies. Visual

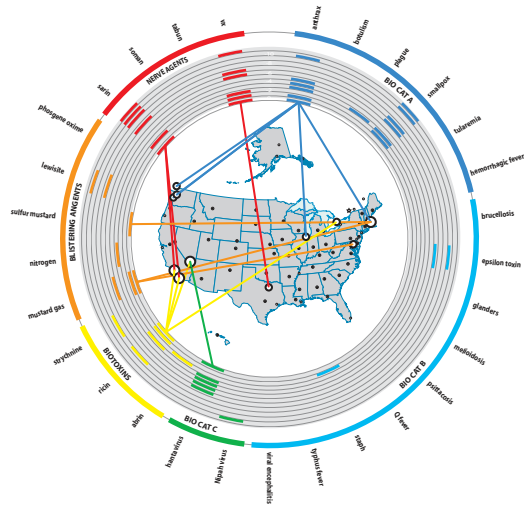


Fig. 5. VisAware for BioWatch (© 2005 IEEE) [14].

analytics tools can not only help revealing the structure of a software package, but can also be used for various other tasks like debugging, maintenance, restructuring or optimization, therefore reducing software maintenance costs. Two applications in this field are CVSScan [17] for interactively tracking the changes of a software package over time, or the Voronoi treemaps [18] for visualization of software metrics.

4.7 Biology, Medicine and Health

The research fields in biology and medicine offer a very wide variety of applications. As computer tomography and ultrasound imaging in the medical area for 3-dimensional digital reconstruction and visualization have been widely used for years, especially the emerging area of bio-informatics now offers a lot of possible applications for visual analytics. From the early beginning of sequencing, scientist in these areas face unprecedented volumes of data, like in the Human Genome Project with three billion base pairs per human. Other new areas like Proteomics (studies of the proteins in a cell), Metabolomics (systematic study of unique chemical fingerprints that specific cellular processes leave behind) or combinatorial chemistry with tens of millions of compounds even enlarge the amount of data every day. A brute-force computation of all possible combinations is often not possible, but visual approaches can help to identify the main regions of interest and exclude areas that are not promising. As traditional visualization techniques can not cope with these amounts of data, new and more effective visualizations are necessary to analyze this amount of data ([19], [20]).

4.8 Engineering Analytics

The application field in engineering analytics covers the whole range from engineering to construction, with a lot of parallels to physics (see above). The most important application is also flow visualization, regarding the automotive industry for example optimization of the air resistance of vehicles, optimization of the flows inside a catalytic converter or diesel particle filter, or computation of optimal air flows inside an engine [21]. Instead of only solving these problems algorithmically, visual analytics can help to understand the flows, and to interactively change construction parameters to optimize the flows. Another application in the automotive industry is the simulation of a car crash, where the frame of a car is represented as a grid of hundreds of thousands of points and the crash is simulated inside a computer. As an optimal car frame cannot be fully automatically computed, visual analytics can help engineers to understand the deformation of the frame during a crash step by step, and to identify the key-points where optimization of the frame is necessary for a better overall stability.

4.9 Personal Information Management

The field of personal information management has many facets and is already affecting our everyday life through digital information devices such as PDAs, mobile phones, and laptop computers. However, there are many further possibilities where research might help to form our future. One example is the IBM Remail project [22], which tries to enhance human capabilities to cope with email overload. Concepts like “Thread Arcs”, “Correspondents Map”, and “Message Map” support the user to efficiently analyse his personal email communication. MIT’s project Oxygen [23] goes one step further, by addressing the challenges of new systems to be pervasive, embedded, nomadic, adaptable, powerful, intentional and eternal. Many of those challenges reflect the visual analytics approach to combine human intelligence and intuition with computational resources.

4.10 Mobile Graphics and Traffic

As an example for traffic monitoring, we consider an ongoing project at University of Illinois-Chicago National Center for Data Mining [24]. In this project, traffic data from the tri-state region (Illinois, Indiana, and Wisconsin) are collected from hundreds of embedded sensors. The sensors are able to identify vehicle weights and traffic volumes. There are also cameras that capture live video feeds, Global Positioning System (GPS) information from selected vehicles, textual accident reports, and weather information. The research challenge is to integrate this massive information flow, provide visualizations that fuse this information to show the current state of the traffic network, and develop algorithms that will detect changes in the flows. Part of this project will involve characterizing normal and expected traffic patterns and developing models that will predict traffic activity when stimulus to the network occurs. The changes detected will include both changes in current congestion levels and differences in congestion levels from what would be expected from normal traffic levels.

5 Technical Challenges

To complete the list of challenges of the previous section, we briefly list the 10 most important technical challenges.

The first technical challenge lies in the field of *problem solving, decision science, and human information discourse*. The process of problem solving supported by technology requires understanding of technology on the one hand, but also comprehension of logic, reasoning, and common sense on the other hand. Intuitive displays and interaction devices should be constructed to communicate analytical results through meaningful visualizations and clear representations.

User acceptability is a further challenge; many novel visualization techniques have been presented, yet their wide-spread deployment has not taken place, primarily due to the users' refusal to change their working routines. Therefore, the advantages of visual analytics tools need to be communicated to the audience of future users to overcome usage barriers, and to eventually tap the full potential of the visual analytics approach. After having developed a system, its *evaluation* is crucial for future reference. Clear comparisons with previous systems to assess its adequacy and objective rules of thumbs to facilitate design decisions would be a great contribution to the community.

To automatically derive *semantics* out of large document collections is still a challenge. On the one hand, some expert systems have been successfully built for specialized fields, but on the other hand the researched methods only perform reasonably within a limited scope. Unlike human comprehension, automatic methods often fail to recognize complex coherences for which they have not been explicitly trained. Modeling of semantics to better deal with conflicting and incomplete information is therefore a challenging field.

Data quality and uncertainty is an issue in many domains, ranging from terrorism informatics to natural sciences, and needs to be taken into account when designing algorithms and visualization metaphors. Semiotic misinterpretations can occur easily. *Data provenance* as the science of understanding where data has come from and why it arrived in the user's database [25] is closely connected to the latter topic. In application fields such as biology, experimental data is made publicly available on the web, copied into other databases, and transformed several times (data curation). Seldom, this information about the transformations and the origins of the data under consideration is properly documented, although it is indispensable for the reproducibility of scientific results. Another challenge lies in *data streams* producing new data at astonishing pace. In this field, especially the timely analysis of the data streams plays an important role. In many cases, e.g. network traffic monitoring, detailed information is abundant and in the long term storage capacities do not suffice to log all data. Thus, efficient and effective methods for compression and feature extraction are needed.

Due to improved measurement methods and decreasing costs of storage capacities, data sets keep on growing. Eventually, *scalability* becomes a major problem in both, automatic as well as visual analysis ([26], [27]), as it becomes more and more challenging to analyze these data sets. For more details see [1], page 24ff "The Scalability Challenge".

Real-world applications often consist of a series of heterogeneous problems. While solving one or more of these problems might still be accomplishable, their correlation make it very difficult to solve the overall problem, thus turning *synthesis of problems* into another challenge. It soon becomes apparent that *integration* with automated analysis, databases, statistics, design, perception, etc. comprises the last of the technical challenges.

6 Conclusion

Visual analytics is an emerging field of research combining strengths from information analytics, geospatial analytics, scientific analytics, statistical analytics, knowledge discovery, data management & knowledge representation, presentation, production and dissemination, cognition, perception and interaction. Its goal is to gain insight into homogeneous, contradictory and incomplete data through the combination of automatic analysis methods with human background knowledge and intuition.

In this paper, we defined the scope of this emerging field and took a closer look at the visual analytics process. By presenting a formal model of the process, we identified the key concepts (data sets, hypotheses, visualizations and insight) and transition functions from one concept to another. To represent the iterative character of the process, a feedback-loop was introduced starting the process over again. To better understand the new analysis methodology, we presented the visual analytics mantra “analyse first - show the important - zoom, filter and analyse further - details on demand”. By means of the top 10 application challenges and the top 10 technical challenges, we gave an overview of the current state of the field and its challenges.

References

1. J. Thomas and K. Cook, *Illuminating the Path: Research and Development Agenda for Visual Analytics*. IEEE-Press, 2005.
2. C. Ware, *Information Visualization - Perception for Design*, 1st ed. Morgan Kaufmann Publishers, 2000.
3. J. W. Tuckey, *Exploratory Data Analysis*. Addison-Wesley, Reading MA, 1977.
4. J. J. Thomas and K. A. Cook, “A Visual Analytics Agenda,” *IEEE Transactions on Computer Graphics and Applications*, vol. 26, no. 1, pp. 12–19, January/February 2006.
5. P. C. Wong and J. Thomas, “Visual analytics,” *IEEE Computer Graphics and Applications*, vol. 24, no. 5, pp. 20–21, 2004.
6. J. J. van Wijk, “The value of visualization,” in *IEEE Visualization*, 2005, pp. 79–86.
7. B. Shneiderman, “The eyes have it: A task by data type taxonomy for information visualizations,” in *IEEE Symposium on Visual Languages*, 1996, pp. 336–343.
8. D. A. Keim, J. Kohlhammer, and J. Thomas. (2005) Workshop on visual analytics. http://infovis.uni-konstanz.de/events/ws-visual_analytics.05/.

9. K.-L. Ma, E. Lum, H. Yu, H. Akiba, M.-Y. Huang, Y. Wang, and G. Schussman, "Scientific discovery through advanced visualization," in *Proceedings of DOE Sci-DAC 2005 Conference*, San Francisco, June 2005.
10. (2007) Sloan Digital Sky Survey. <http://www.sdss.org/>.
11. (2007) COMPLETE - the COordinated Molecular Probe Line Ex-tinction Thermal Emission survey of star forming regions. <http://cfa-www.harvard.edu/COMPLETE/index.html>.
12. D. A. Keim, T. Nietzschmann, N. Schelwies, J. Schneidewind, T. Schreck, and H. Ziegler, "FinDEx: A spectral visualization system for analyzing financial time series data," in *EuroVis 2006: Eurographics/IEEE-VGTC Symposium on Visualization, Lisbon, Portugal, 8-10 May, 2006*.
13. M. Wattenberg, "Visualizing the stock market," in *CHI '99: CHI '99 extended abstracts on Human factors in computing systems*. New York, NY, USA: ACM Press, 1999, pp. 188–189.
14. Y. Livnat, J. Agutter, S. Moon, and S. Foresti, "Visual correlation for situational awareness," in *IEEE Symposium on Information Visualization, 2005*, pp. 95–102.
15. S. T. Teoh, T. Jankun-Kelly, K.-L. Ma, and S. F. Wu, "Visual data analysis for detecting flaws and intruders in computer network systems," *IEEE Transactions on Computer Graphics and Applications*, pp. 27–35, September/October 2004.
16. J. R. Goodall, W. G. Lutters, P. Rheingans, and A. Komlodi, "Preserving the big picture: Visual network traffic analysis with TNV," in *Proceedings of IEEE Workshop on Visualization for Computer Security, 2005*, pp. 47–54.
17. S. Voinea and A. T. M. Chaudron, "Version-centric visualization of code evolution," in *Proceedings of Eurographics/IEEE-VGTC Symposium on Visualization, 2005*.
18. M. Balzer and O. Deussen, "Voronoi treemaps," in *IEEE Symposium on Information Visualization (InfoVis 2005)*, 2005, pp. 7–14.
19. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal on Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.
20. T. Tatusova and T. Madden, "Blast2 sequences - a new tool for comparing protein and nucleotide sequences," *FEMS Microbiology Letter*, vol. 174, no. 2, pp. 247–250, 1999.
21. H. Doleisch, M. Mayer, M. Gasser, R. Wanker, and H. Hauser, "Case study: Visual analysis of complex, time-dependent simulation results of a diesel exhaust system," in *6th Joint IEEE TCVG EUROGRAPHICS Symposium on Visualization (VisSym 2004)*, May 2004, pp. 91–96.
22. (2005) IBM Remail - reinventing email. <http://www.research.ibm.com/remail/>.
23. (2007) MIT Project Oxygen. <http://oxygen.lcs.mit.edu/>.
24. (2007) Pantheon Highway Gateway. <http://highway.lac.uic.edu/>.
25. P. Buneman, S. Khanna, and W.-C. Tan, "Why and where: A characterization of data provenance," in *Database Theory - ICDT 2001: 8th International Conference, London, UK, January 2001*, ser. Lecture Notes in Computer Science, J. V. den Bussche and V. Vianu, Eds., vol. 1973. Springer, Jan 2001, p. 316.
26. C. Chen, "Top 10 unsolved information visualization problems," *IEEE Transactions on Computer Graphics and Applications*, vol. 25, no. 4, pp. 12–19, July/August 2005.
27. S. G. Eick and A. F. Karr, "Visual scalability," *Journal of Computational & Graphical Statistics*, pp. 22–43, March 2002.